Fast calibration of weak FARIMA models

Youssef ESSTAFA

25 August 2023

Le Mans Université,
Laboratoire Manceau de Mathématiques,
Institut du Risque et de l'Assurance.

Joint work with **Samir BEN HARIZ**, **Alexandre BROUSTE** and **Marius SOLTANE**

Outline

Introduction

Long memory processes

- Let $(X_t)_{t\in\mathbb{Z}}$ be a second order stationary process.

- Denote by $\gamma_X(\cdot)$ its autocovariance function and by $\rho_X(\cdot)$ its autocorrelation function, *i.e.* $\forall t, h \in \mathbb{Z}$,

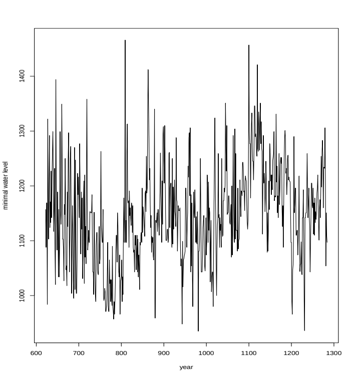
$$\gamma_X(h) = \mathrm{Cov}(X_t, X_{t+h}) \quad \text{and} \quad \rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}.$$

Definition 1.
The process $(X_t)_{t\in\mathbb{Z}}$ is called a long memory process, in the covariance sense, if

$$\sum_{h=-\infty}^{\infty} |\gamma_X(h)| = \infty.$$

An illustrative example

Nile River Minima



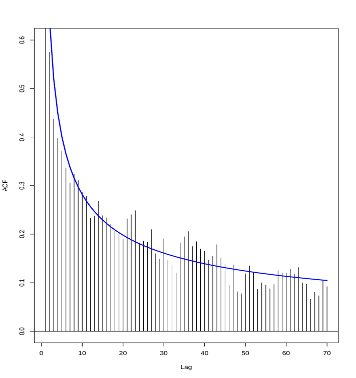


Figure 1: Annual minima of the water level in the Nile river for the years 622 to 1281.

Figure 2: Empirical autocorrelations of the Nile Water Minima series. The curve in blue is that of $x \to 0.91/x^{\mathbf{0.51}}$.

Examples of long memory processes

Example 1: fractional Gaussian noise, [Mandelbrot and Wallis (1969)].

Definition 2: fBm, [Mandelbrot and Van Ness (1968)].

The fractional Brownian motion with Hurst exponent $0 < H < 1$, denoted $(B_{\mathrm{H}}(t))_{t\in\mathbb{R}}$, is the unique continuous centered Gaussian process whose covariance is given by

$$\mathbb{E}\left[B_{\mathrm{H}}(t)B_{\mathrm{H}}(s)\right] = \frac{\sigma_{\mathrm{H}}^2}{2}\left(|t|^{2\mathrm{H}} + |s|^{2\mathrm{H}} - |t-s|^{2\mathrm{H}}\right),$$

where $\sigma_{\mathrm{H}}^2 = \mathrm{Var}\{B_{\mathrm{H}}(1)\}$.

The fractional Gaussian noise $(\epsilon_t^{\mathrm{H}})_{t\in\mathbb{Z}}$ is the increment process of the fractional Brownian motion $(B_{\mathrm{H}}(t))_{t\in\mathbb{R}}$, *i.e.* $\forall t \in \mathbb{Z}$,

$$\epsilon_t^{\mathrm{H}} = B_{\mathrm{H}}(t+1) - B_{\mathrm{H}}(t).$$

- Using the structure of the autocovariance function of $(B_{\mathrm{H}}(t))_{t\in\mathbb{R}}$, we deduce that for all $k \in \mathbb{Z}$,

$$\gamma_{\epsilon^{\mathrm{H}}}(k) := \mathrm{Cov}(\epsilon_1^{\mathrm{H}}, \epsilon_{1+k}^{\mathrm{H}}) = \frac{\sigma_{\mathrm{H}}^2}{2}\left(|k+1|^{2\mathrm{H}} + |k-1|^{2\mathrm{H}} - 2|k|^{2\mathrm{H}}\right).$$

- A Taylor expansion of $\ell : x \to (1-x)^{2\mathrm{H}} - 2 + (1+x)^{2\mathrm{H}}$ at 0 implies that for sufficiently large k,

$$\gamma_{\epsilon^{\mathrm{H}}}(k) = \frac{\sigma_{\mathrm{H}}^2}{2} k^{2\mathrm{H}} \ell(1/k) = \sigma_{\mathrm{H}}^2 \mathrm{H}(2\mathrm{H}-1) k^{2\mathrm{H}-2} + \mathrm{o}(k^{2\mathrm{H}-2}).$$

Conclusion.

- When $0 < \mathrm{H} < 1/2$, the process $(\epsilon_t^{\mathrm{H}})_{t\in\mathbb{Z}}$ is anti-persistent.
- If $\mathrm{H} = 1/2$, $(\epsilon_t^{\mathrm{H}})_{t\in\mathbb{Z}}$ is an independent process.
- In the case where $1/2 < \mathrm{H} < 1$, $(\epsilon_t^{\mathrm{H}})_{t\in\mathbb{Z}}$ is a long memory process.

Definition 3.

A process $(X_t)_{t\in\mathbb{Z}}$ is an ARMA(p, q) if $(X_t)_{t\in\mathbb{Z}}$ is stationary and satisfies

$$X_t = a_1 X_{t-1} + \cdots + a_p X_{t-p} + \epsilon_t - b_1 \epsilon_{t-1} - \cdots - b_q \epsilon_{t-q}, \tag{1}$$

where $a_1, \ldots, a_p, b_1, \ldots, b_q \in \mathbb{R}$ and $(\epsilon_t)_{t\in\mathbb{Z}}$ is a white noise.

Equation (1) can be rewritten in the compact form

$$a(L)X_t = b(L)\epsilon_t,$$ [1]

where

$$a(z) = 1 - \sum_{i=1}^{p} a_i z^i \quad \text{and} \quad b(z) = 1 - \sum_{i=1}^{q} b_i z^i.$$

Problem!

The autocovariance function of the process defined in (1) satisfies $\gamma_X(h) \sim C\rho^h$ ($C \neq 0$ and $0 < \rho < 1$), when $h \to \infty$.

[1] In this equation, L stands for the back-shift operator, *i.e.* for any non-negative integer k, $L^k X_t = X_{t-k}$.

Examples of long memory processes

Example 2: FARIMA[2] models, [Granger and Joyeux (1980), Hosking (1981)].

Definition 4: FARIMA processes.

A process $(X_t)_{t\in\mathbb{Z}}$ is said to be a FARIMA(p,d,q) with $d \in]-0.5, 0.5[$ if $(X_t)_{t\in\mathbb{Z}}$ is stationary and satisfies the difference equations,

$$a(L)(1-L)^d X_t = b(L)\epsilon_t, \tag{2}$$

where L is the back-shift operator, $(\epsilon_t)_{t\in\mathbb{Z}}$ is a white noise and $a(\cdot)$, $b(\cdot)$ are polynomials of degrees p, q respectively.

The fractional difference operator $(1-L)^d$ is given by

$$(1-L)^d = \sum_{j=0}^{+\infty} \alpha_j(d) L^j, \text{ where } \alpha_j(d) = \frac{d(d-1)\cdots(d-j+1)}{j!}(-1)^j.$$

[2] Fractional AutoRegressive Integrated Moving Average.

Least squares estimation of weak FARIMA models

Weak FARIMA processes

Let $(X_t)_{t\in\mathbb{Z}}$ be a second-order stationary process.

Definition 5: Weak FARIMA processes.

The process $(X_t)_{t\in\mathbb{Z}}$ is a weak FARIMA(p, d_0, q) process if it satisfies (2) with $d_0 \in]0, 1/2[$[3] and if the innovations process $(\epsilon_t)_{t\in\mathbb{Z}}$ is a weak white noise[4] of variance $\sigma_\epsilon^2 > 0$.

Remarks.

In a weak FARIMA:

- No constraints on the distribution of $(\epsilon_t)_{t\in\mathbb{Z}}$[5].
- The process $(\epsilon_t)_{t\in\mathbb{Z}}$ may contain very general nonlinear dependencies.

[3] The process is long memory in this case.

[4] Weak white noise is a centered, uncorrelated process with finite variance.

[5] We adopt here a semi-parametric approach for estimating weak FARIMA models.

Least squares estimator (LSE)

Theoretical frame:

Let $\tilde{\Theta}$ be the parameter space

$$\tilde{\Theta} := \{\tilde{\theta} = (\theta_1, \theta_2, \ldots, \theta_{p+q})' \in \mathbb{R}^{p+q};\ a_{\tilde{\theta}}(z) = 1 + \theta_1 z + \cdots + \theta_p z^p \text{ and } b_{\tilde{\theta}}(z) = 1 + \theta_{p+1} z + \cdots + \theta_{p+q} z^q \text{ have all their zeros outside the unit disk}\}.$$

- Denote by Θ the Cartesian product $\tilde{\Theta} \times [d_1, d_2]$, where $[d_1, d_2] \subset]0, 1/2[$ with $d_1 \leq d_0 \leq d_2$.
- The parameter of interest $\theta_0 := (a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_q, d_0)'$ is supposed to belong to the parameter space Θ.
- For all $\theta = (\tilde{\theta}', d)' \in \Theta$, we define $(\epsilon_t(\theta))_{t \in \mathbb{Z}}$ as the stationary process which is the solution of

$$\epsilon_t(\theta) = \sum_{j \geq 0} \alpha_j(d) X_{t-j} + \sum_{i=1}^{p} \theta_i \sum_{j \geq 0} \alpha_j(d) X_{t-i-j} - \sum_{j=1}^{q} \theta_{p+j} \epsilon_{t-j}(\theta).$$

Least squares estimator (LSE)

Given a realization $X_1, X_2, \ldots, X_n$ of length n, $\epsilon_t(\theta)$ can be approximated, for $0 < t \leq n$, by $\tilde{\epsilon}_t(\theta)$ defined recursively by

$$\tilde{\epsilon}_t(\theta) = \sum_{j=0}^{t-1} \alpha_j(d) X_{t-j} + \sum_{i=1}^{p} \theta_i \sum_{j=0}^{t-i-1} \alpha_j(d) X_{t-i-j} - \sum_{j=1}^{q} \theta_{p+j} \tilde{\epsilon}_{t-j}(\theta),$$

with $\tilde{\epsilon}_t(\theta) = X_t = 0$ if $t \leq 0$.

Lemma 1.

These initial values are asymptotically negligible uniformly in θ. More precisely, if $(\epsilon_t)_{t \in \mathbb{Z}}$ is strictly stationary and ergodic,

$$\lim_{t \to \infty} \sup_{\theta \in \Theta} |\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)| = 0 \quad \text{a.s.}$$

The random variable $\hat{\theta}_n$ is called least squares estimator if it satisfies, almost surely,

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}}\ Q_n(\theta), \text{ where } Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \tilde{\epsilon}_t^2(\theta).$$

Asymptotic properties of the LSE

$\rightarrow$ Our first two main results concern the strong consistency and the asymptotic normality of the LSE of the parameter θ_0.

$\rightarrow$ The strong consistency of the LSE is obtained under the following assumption:

(A1): The process $(\epsilon_t)_{t\in\mathbb{Z}}$ is strictly stationary and ergodic.

Theorem 1 (strong consistency).

Assume that $(\epsilon_t)_{t\in\mathbb{Z}}$ satisfies (2). Let $(\hat{\theta}_n)_{n\geq 1}$ be a sequence of least squares estimators. Under **(A1)**, we have

$$\hat{\theta}_n \xrightarrow[n\to\infty]{\text{a.s.}} \theta_0.$$

Strong mixing coefficients

The strong mixing coefficients $\{\alpha_\epsilon(h)\}_{h\geq 0}$ of the process $(\epsilon_t)_{t\in\mathbb{Z}}$ are defined by

$$\alpha_\epsilon(h) = \sup_{A\in\mathcal{F}_{-\infty}^t, B\in\mathcal{F}_{t+h}^\infty} \left|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\right|,$$

where $\mathcal{F}_{-\infty}^t = \sigma(\epsilon_u, u\leq t)$ and $\mathcal{F}_{t+h}^\infty = \sigma(\epsilon_u, u\geq t+h)$.

Consider

(A2): There exists an integer τ such that for some $\nu\in(0,1]$, we have $\mathbb{E}|\epsilon_t|^{\tau+\nu} < \infty$ and $\sum_{h=0}^\infty (h+1)^{k-2}\left\{\alpha_\epsilon(h)\right\}^{\nu/(k+\nu)} < \infty$ for $k = 1,\dots,\tau$.[6]

[6] See [Doukhan and León (1989)].

Asymptotic properties of the LSE

The asymptotic normality of the LSE is stated in the following theorem:

Theorem 2 (asymptotic normality).

Assume that $(\epsilon_t)_{t\in\mathbb{Z}}$ satisfies (2) and $\theta_0 \in \overset{\circ}{\Theta}$. Let $(\hat{\theta}_n)_{n\geq 1}$ be a sequence of least squares estimators. Under **(A1)** and **(A2)** with $\tau = 4$, the sequence

$$\left(\sqrt{n}(\hat{\theta}_n - \theta_0)\right)_{n\geq 1}$$

has a limiting centered normal distribution with covariance matrix $\Omega := J^{-1}IJ^{-1}$, where

$$I = \lim_{n\to\infty} \mathrm{Var}\left\{\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0)\right\} \text{ and } J = \lim_{n\to\infty}\left\{\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta_0)\right\} \text{ a.s.}$$

Using the stationarity of $(H_t)_{t\in\mathbb{Z}}$, defined by $H_t = 2\epsilon_t\{\partial\epsilon_t(\theta_0)/\partial\theta\}$, and **(A2)** with $\tau = 4$, we show that

$$J = 2\mathbb{E}\left[\frac{\partial}{\partial\theta}\epsilon_t(\theta_0)\frac{\partial}{\partial\theta'}\epsilon_t(\theta_0)\right] \text{ and } I = \sum_{h=-\infty}^{\infty} \mathrm{Cov}\left(H_t, H_{t-h}\right).$$

Remarks.

- The matrix J has the same expression in the strong[7] and weak FARIMA cases.
- In the standard strong FARIMA case, we have

$$I = 2\sigma_\epsilon^2 J.$$

  Thus, the asymptotic covariance matrix of the LSE is then reduced as $\Omega_S := 2\sigma_\epsilon^2 J^{-1}$.

[7] In this case, the noise $(\epsilon_t)_{t\in\mathbb{Z}}$ is assumed to be an iid sequence of random variables.

Le Cam's one-step estimation of weak FARIMA models

One-step estimator

For $n \geq 1$ and $\theta \in \Theta$, recall that our objective function is given by

$$Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \tilde{\epsilon}_t^2(\theta), \tag{3}$$

where $(\tilde{\epsilon}_t(\theta))_{t\in\mathbb{Z}}$ is the observable noise process.

The Le Cam one-step estimator is defined, almost-surely, by

$$\overline{\theta}_n = \theta_n^* - \left\{ \frac{\partial^2}{\partial\theta\partial\theta'} Q_n\left(\theta_n^*\right) \right\}^{-1} \frac{\partial}{\partial\theta} Q_n\left(\theta_n^*\right), \tag{4}$$

where θ_n^* is the least squares estimator of parameter θ_0 calculated over the first $m = [n^\delta]$, with $0 < \delta \leq 1$, observations $X_1, \ldots, X_m$, *i.e.*

$$\theta_n^* = \underset{\theta\in\Theta}{\operatorname{argmin}}\ Q_m(\theta), \text{ where } Q_m(\theta) = \frac{1}{[n^\delta]} \sum_{t=1}^{[n^\delta]} \tilde{\epsilon}_t^2(\theta). \tag{5}$$

One-step estimator

Remarks.

- The consideration of the initial LSE on a subsample of size $m = [n^\delta]$ greatly reduces the computation time for the estimation of the parameters in the model.
- For $\delta > 1/2$, a sole Fisher scoring correcting step is sufficient to reach similar asymptotic properties as the LSE on the whole sample.
- If $\delta \leq 1/2$, the one-step estimator remains consistent but similar asymptotic normality as the LSE requires multiple Fisher scoring steps.

Asymptotic properties of the OS estimator

Under the same assumptions as those considered for the least squares estimator, we show:

Theorem 3 (strong consistency).

Assume that $(\epsilon_t)_{t\in\mathbb{Z}}$ satisfies (2). Let $(\overline{\theta}_n)_{n\geq 1}$ be the sequence of Le Cam's one-step estimators defined by (4). Under Assumption **(A1)**, we have

$$\overline{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0.$$

Theorem 4 (asymptotic normality).

Assume that $(\epsilon_t)_{t\in\mathbb{Z}}$ satisfies (2) and $\theta_0 \in \overset{\circ}{\Theta}$. Under **(A1)** and **(A2)** with $\tau = 4$, the sequence $\{\sqrt{n}(\overline{\theta}_n - \theta_0)\}_{n\geq 1}$ with $\delta > 1/2$ has a limiting centered normal distribution with covariance matrix $\Omega := J^{-1} I J^{-1}$.

Sketch of the proof of asymptotic normality

Proposition (stochastic Lipschitz property).

Assume that $(X_t)_{t\in\mathbb{Z}}$ satisfies (2). For any $i, j \in \{1, \ldots, p+q+1\}$ and all $\theta^{(1)}, \theta^{(2)} \in \Theta$, one has

$$\left| \frac{\partial^2}{\partial\theta_i \partial\theta_j} Q_n\left(\theta^{(1)}\right) - \frac{\partial^2}{\partial\theta_i \partial\theta_j} Q_n\left(\theta^{(2)}\right) \right| \leq \Delta_n \left\| \theta^{(1)} - \theta^{(2)} \right\|,$$

where Δ_n is bounded in probability.

In view of (4) and by Taylor expansion of the function $\partial Q_n(\cdot)/\partial\theta$ around θ_0, we have

$$\sqrt{n}\left(\overline{\theta}_n - \theta_0\right) = \sqrt{n}\left(\theta_n^* - \theta_0\right) - \sqrt{n}\left\{\frac{\partial^2}{\partial\theta\partial\theta'} Q_n\left(\theta_n^*\right)\right\}^{-1} \times \left\{\frac{\partial}{\partial\theta} Q_n\left(\theta_0\right) + \left[\frac{\partial^2}{\partial\theta_i\partial\theta_j} Q_n\left(\tilde{\theta}_{n,i,j}\right)\right]\left(\theta_n^* - \theta_0\right)\right\},$$

where the $\tilde{\theta}_{n,i,j}$'s are between θ_n^* and θ_0.

Hence, it follows that

$$\begin{aligned}
&\sqrt{n}\left(\overline{\theta}_n - \theta_0\right) \\
&\quad = \left\{\frac{\partial^2}{\partial\theta\partial\theta'} Q_n\left(\theta_n^*\right)\right\}^{-1} n^{\delta/2}\left\{\frac{\partial^2}{\partial\theta\partial\theta'} Q_n\left(\theta_n^*\right) - \left[\frac{\partial^2}{\partial\theta_i\partial\theta_j} Q_n\left(\tilde{\theta}_{n,i,j}\right)\right]\right\} \\
&\qquad \times n^{\delta/2}\left(\theta_n^* - \theta_0\right) n^{1/2-\delta} - \left\{\frac{\partial^2}{\partial\theta\partial\theta'} Q_n\left(\theta_n^*\right)\right\}^{-1}\sqrt{n}\frac{\partial}{\partial\theta} Q_n\left(\theta_0\right).
\end{aligned} \tag{6}$$

- The second term on the rhs of (6) converges in law to $\mathcal{N}(0, J^{-1} I J^{-1})$.

- The first term converges in probability to 0. In fact:

→ The quantity $n^{\delta/2}(\theta_n^* - \theta_0) = \mathrm{O}_{\mathbb{P}}(1)$ due to the $n^{\delta/2}-$consistency of the initial estimator.

→ The matrix $n^{\delta/2}\{\frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta_n^*) - [\frac{\partial^2}{\partial\theta_i\partial\theta_j} Q_n(\tilde{\theta}_{n,i,j})]\} = \mathrm{O}_{\mathbb{P}}(1)$ due to the proposition before.

Some simulations

$\rightarrow$ We numerically study the behavior of the LSE and the Le Cam one-step estimator of the memory parameter for FARIMA models of the form

$$(1-L)^d (X_t + aX_{t-1}) = \epsilon_t + b\epsilon_{t-1}, \tag{7}$$

where $(a, b, d) = (0.2, 0.5, 0.3)$.

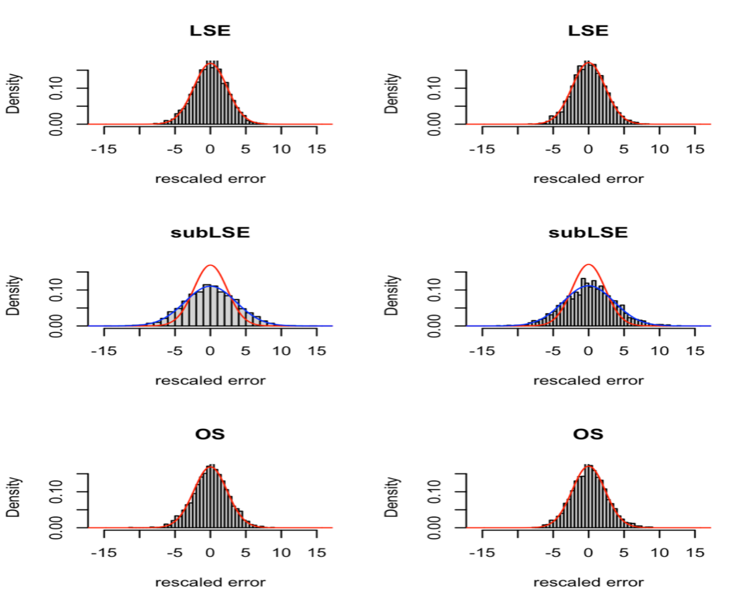
$\rightarrow$ We consider the following two cases:

- The process $(\epsilon_t)_t$ is a centered iid Gaussian process with variance 1.
- The innovation process in (7) is defined, for all $t \in \mathbb{Z}$, by

$$\epsilon_t = \eta_t^2 \eta_{t-1}, \tag{8}$$

  where $(\eta_t)_t$ is an iid $\mathcal{N}(0,1)$ process.

$\rightarrow$ We simulated $M = 2,000$ independent trajectories of size $n = 5,000$ of (7) endowed first by the strong noise and then by the weak noise (8). We consider that $\delta = 0.9$.


LSE
Density
0.00 0.10
-15 -5 0 5 10 15
rescaled error
LSE
Density
0.00 0.10
-15 -5 0 5 10 15
rescaled error
subLSE
Density
0.00 0.10
-15 -5 0 5 10 15
rescaled error
subLSE
Density
0.00 0.10
-15 -5 0 5 10 15
rescaled error
OS
Density
0.00 0.10
-15 -5 0 5 10 15
rescaled error
OS
Density
0.00 0.10
-15 -5 0 5 10 15
rescaled error

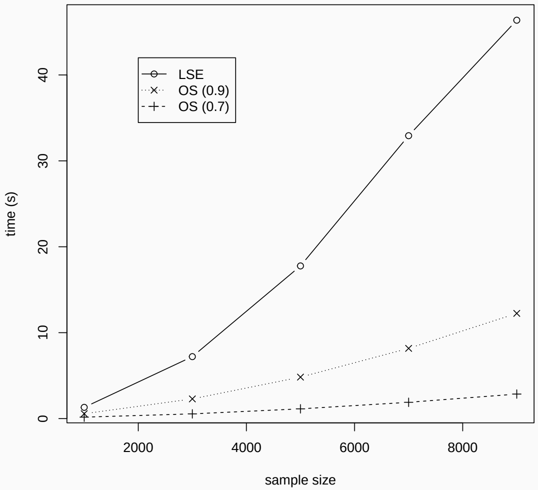


Figure 3: Comparison of the computation times with respect to the sample size of the LSE and the one-step estimators of the parameters of Model (7) induced by Noise (8). For each size n, 1,000 replications are generated.

Thank you for your attention